GOVERNMENTS DIVISION REPORT SERIES
(Research Report #2010-3)

Variance Estimation for Decision-Based Estimators with Application to the Annual Survey of Public Employment and Payroll

Yang Cheng
Eric Slud
Carma Hogue

U.S. Census Bureau
Washington, DC 20233

CITATION: Cheng, Yang, Eric Slud, Carma Hogue. 2010. Variance Estimation for Decision-Based Estimators with Application to the Annual Survey of Public Employment and Payroll. Governments Division Report Series, Research Report #2010-3

Report Completed: September 23, 2010
Report Issued: October 7, 2010

Variance Estimation for Decision-Based Estimators with Application to the Annual Survey of Public Employment and Payroll

Yang Cheng, Eric Slud, Carma Hogue
Governments Division and Statistical Research Division
U.S. Census Bureau[1], Washington, D.C. 20233-0001

Abstract: In 2009, two major surveys in the Governments Division of the Census Bureau were redesigned to reduce sample size, save resources, and improve the precision of the estimates. We developed a new decision-based estimation method in which the collapsing of strata either by state and government type or by small and large size was determined by a series of hypothesis tests of the equality of fitted coefficients in linear relationships between attributes and their values in the previous census year. In this research, we study design-based variance estimation by a bootstrap method, for the new decision-based stratified regression estimates applied to the Annual Survey of Public Employment and Payroll. The bootstrap method, which goes beyond available theory, is validated through a small simulation study. We use the data from the 2007 Census of Governments Employment to illustrate our methods.

Key Words: survey design, decision-based estimation, re-sampling method, bootstrap, mean squared error

1. Introduction

The Annual Survey of Public Employment and Payroll (ASPEP) provides current estimates for full-time and part-time state and local government employment and payroll by government function (i.e., elementary and secondary education, higher education, police protection, fire protection, financial administration, judicial and legal, etc.). This survey covers all state and local governments in the United States, which include counties, cities, townships, special districts, and school districts. The first three types of governments are referred to as general-purpose governments as they generally provide several governmental functions. School districts cover only the education function. Special districts usually provide one function code, but can provide more than one function. ASPEP is the only source of public employment data by program function and selected job category. Data on employment include number of full-time and part-time employees, gross pay, and hours paid for part-time employees. Reported data are for the government's pay period that includes March 12. Data collection begins in March and continues for about seven months.

There are 89,526 state and local government units in our universe. In 2009, after exploring possible cut-off sample methods for ASPEP, we developed a new modified cut-off sample method based on the current stratified probability proportional-to-size sample design in order to reduce the sample size, save resources, and improve the precision of the estimates. Additional benefits from using this modified cutoff sample design are to reduce respondent burden, improve data quality, and increase physical efficiency. The modified cut-off sample method is applied in two stages. We first select a state-by-governmental type stratified probability proportional-to-size (PPS) sample. The

---

[1] *Disclaimer: This report is released to inform interested parties of research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.*

PPS sample is based on total payroll, which is the sum of full-time pay and part-time pay, from the Employment portion of the 2007 Census of Government. In the second stage, we construct a cut-off point, to distinguish small and large governmental units in the stratum. Then we sub-sample in the substratum with small-size governmental units.

The design is specified by two parameters: (1) the cut-off point, which determines how to construct size-based substrata; and (2) the small-unit sub-sampling rate, which is the proportion of small government units to be sampled randomly. We know the cut-off must be between the minimum of total payroll in 2007 and the maximum of payroll in 2007, and the reduction rate must be between 0 and 1. The whole design is supported by the following ***Lemma***.

Lemma 1: Suppose a sample S is a probability proportional to size (PPS) sample with sample size n drawn from universe U of known size N. Suppose further that the sub-ample $S_m \subset S$ is to be drawn by simple random sampling taking m out of n. Then, S_m is a PPS sample with size m, and the second-order inclusion probabilities for distinct pairs of elements of the sub-sample are also proportional to the corresponding joint inclusion probabilities for the sample S.

Proof: Let the size measure z_i be known for all elements $i \in U$. Then, the inclusion probability is $\pi_i = P(i \in S) = nz_i \Big/ \sum_{k \in U} z_k$ subject to the constraint that $n \cdot \max_{k \in U} z_k \le \sum_{i \in U} z_i$ and the joint inclusion probability is π_{ij}.

The second stage conditional inclusion probabilities are

$$P(i \in S_m \mid i \in S) = \frac{m}{n} \text{ and } P(i, j \in S_m \mid i, j \in S) = \binom{m}{2} \Big/ \binom{n}{2} \text{ for } i \neq j$$

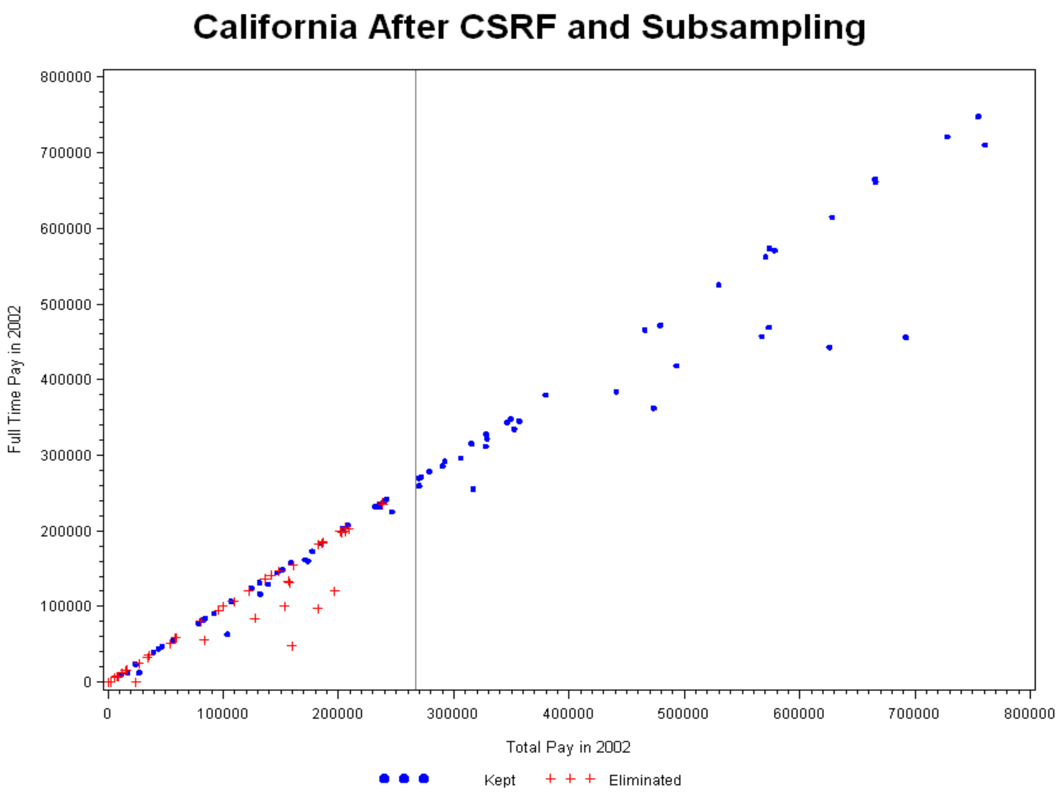
Then the first and second order inclusion probabilities for the sub-sample S_m are given by

$$\pi_i^* = P(i \in S_m) = \frac{m}{n}\pi_i \text{ and } \pi_{ij}^* = P(i, j \in S_m) = \frac{m(m-1)}{n(n-1)}\pi_{ij}$$

Thus, S_m is a PPS sample with joint inclusion probabilities for distinct elements proportional to those in the sample S before sub-sampling.

Figure 1 illustrates how the modified cut-off sampling method works. Using full-time payroll in California Special Districts as an example, we first apply the probability proportional-to-size sample method to select the sample S in **Figure 1.** Second, the size-based substrata are constructed by applying the cumulative square root frequency method to determine the cut-off point with respect to the size of units in the problematic special districts. Thus, we see sub-stratum with small governmental units on the left hand side of the vertical cut-off line and the sub-stratum with large governmental units on the right hand side of the line. Third, we sub-sample in the small-unit sub-stratum. We keep all large governmental units and draw a simple random sample without replacement in the small-unit sub-stratum. The units with red “+” were eliminated from the sample in the second stage.

Figure 1: Illustration of the modified cut-off sample method for special districts in California for full-time payroll versus the size of the government units, which is total pay



Source: U.S. Census Bureau. 2002 Census of Governments: Employment

2. Decision-Based Estimation

We explore the formula for estimating the survey total of key variables: full-time employment, full-time payroll, part-time employment, part-time payroll, and part-time hours. Let Z, the total pay from the most recent (2002) census, be the size variable used in PPS sampling. A general estimation formula for estimating the total is:

$$\hat{t}_y = \sum_{i \in S} w_i y_i \tag{1}$$

where the weights, w_i, may depend on the survey design, attributes, and auxiliary data.

When w_i is a function of survey design only, to be unbiased the estimator (1) must be a Horvitz-Thompson (H-T) estimator with weight $1/\pi_i$, where π_i is the inclusion probability for unit i. Model-assisted estimators of the type (1) can have weights which are functions of survey weights, design-based estimators of frame-population parameters, and auxiliary data. When the regression predictor X is the same variable as Y from the most recent census, the single stratum weighted regression (GREG) estimator is

$$\hat{t}_{y,reg} = \hat{t}_{y,\pi} + \hat{b}(t_x - \hat{t}_{x,\pi}) \qquad (2)$$

where $t_x = \sum_{i \in U} x_i$, $\hat{t}_{x,\pi} = \sum_{i \in S} \frac{x_i}{\pi_i}$, $\hat{t}_{y,\pi} = \sum_{i \in S} \frac{y_i}{\pi_i}$, and $\hat{b} = \frac{\sum_{i \in S} (x_i - \bar{x})(y_i - \bar{y}) / \pi_i}{\sum_{i \in S} (x_i - \bar{x})^2 / \pi_i}$.

The same estimator (2) can also be shown (Sarndal et al. 1992, Deville and Sarndal 1992) to arise as a calibration estimator which uses calibrated weights, chosen as close as possible, according to a weighted-sum-of-squares distance function, to the original sampling design weights $1/\pi_i$ while also respecting a set of constraints, the calibration equations

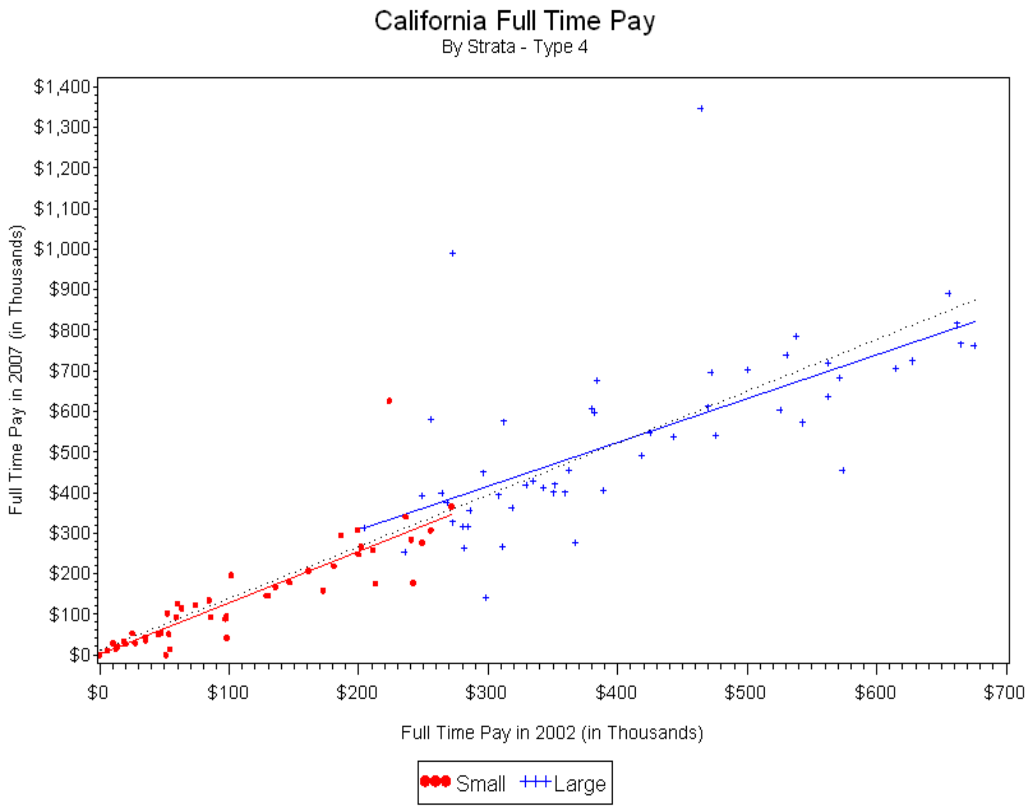
$$\sum_{i \in S} w_i x_i = \sum_{i \in U} x_i \text{ , } \sum_{i \in S} w_i = N \qquad (3)$$

where the values x_i are 'auxiliary', with population total assumed known. That is, the GREG estimator (2) is known to be algebraically identical to (1) if the calibrated weights w_i are defined to minimize $\sum_S \pi_i (w_i - 1/\pi_i)^2$ subject to the constraints (3). Since the calibrated weights in GREG estimation depend on the responses y_i, the GREG estimator (2) is actually nonlinear in the y_i.

Cheng et al. (2009) proposed a decision-based method to improve the precision of estimates and reduce the mean square error of weighted survey total estimates. The idea was to test the equality of linear regression lines to determine whether we can combine data in different substrata. In Cheng et al. (2009), equality of regression lines is tested in two steps. First, a test is performed of the null hypothesis that the slopes are identical. If the P value is less than 0.05, the null hypothesis is rejected in favor of the conclusion that the regression lines are significantly different. In this case, there is no reason to compare the intercepts. If the P value for comparing slopes is greater than 0.05, the null hypothesis of equality of slopes cannot be rejected, but intercepts can be compared. If the regression lines for the two substrata are not found to be significantly different, then a single line is estimated from the combined substrata.

One might ask whether two substratum regression lines with roughly equal slopes might actually be different in our context, that is, might be parallel rather than identical. To examine this possibility, we estimated the slopes and intercepts for substratum data sets in selected state-by-type strata, using 2002 and 2007 Census data. Our data analyses led us to observe that we never rejected the null hypothesis of equality of intercepts when we could not reject the null hypothesis of equality of slopes. This makes sense because the 2007 payrolls can be 0 essentially only if the 2002 payrolls are. Thus, we decided to perform our hypothesis tests for equality of substratum regression lines strictly by testing equality of slopes.

Figure 2: Linear fits for small and large special districts in California for full-time payroll versus a single linear fit for data combining small and large special districts



Source: U.S. Census Bureau. 2002 and 2007 Census of Governments: Employment

Figure 2 displays how the decision-based approach works on small government units as compared with large government units. We use full-time payroll in California as an example. The two solid straight lines are linear regression fits for small and large special districts. They are not the same, but have very similar slopes and a small difference between two intercepts. Since we cannot reject the null hypothesis of equality of the model coefficients and claim model coefficients are significantly different, we combine the small and large government units to reduce model error when we apply the model fit for a combined stratum including both small and large units. The dotted line is the best linear fit for the combined stratum.

Now, we test the null hypothesis $H_0 : b_1 = b_2$, that is, the equality of the frame population regression slopes for two substrata. From equation (2), the model-assisted slope estimator, $\widehat{b}$, can be expressed within each stratum using the PPS design weights as

$$\widehat{b} = \sum_{i \in S} \frac{1}{\pi_i} y_i (x_i - \widehat{t}_{x,\pi} / \widehat{N}) \Big/ \sum_{i \in S} \frac{1}{\pi_i} (x_i - \widehat{t}_{x,\pi} / \widehat{N})^2 \qquad (4)$$

where $\widehat{N} = \widehat{t}_{1,\pi} = \sum_{i \in S} \frac{1}{\pi_i}$. In large samples, $\widehat{b}$ is approximately normally distributed with mean b and a theoretical variance denoted $\sum$. Under the null hypothesis, starting from two sub-strata U_1, U_2 with samples S_1, S_2 of sizes n_1, n_2 and slope estimates $\widehat{b}_1, \widehat{b}_2$, we have $\hat{b}_1 - \hat{b}_2 \sim N(0, \Sigma_{1,2})$, where $\Sigma_{1,2} = \Sigma_1 + \Sigma_2$. Therefore, the test statistic becomes

$$(\hat{b}_1 - \hat{b}_2)\Sigma_{1,2}^{-1}(\hat{b}_1 - \hat{b}_2) \sim \chi_1^2 \qquad (5)$$

We will discuss the variance estimator for $\hat{b}$ in the next section. The critical value for a test based on (5) is obtained from chi-squared percentage points with 1 degree of freedom. If the P value is less than 0.05, we reject the null hypothesis and conclude that the substratum population regression slopes are significantly different.

Combining the hypothesis test and GREG estimator, we can explicitly express the decision-based estimator formula based on sampled data in two substrata in terms of the outcome of the hypothesis test of $H_0 : b_1 = b_2$. If we cannot reject the null hypothesis, then the slopes estimated in S_1 and S_2 are accepted as the same, and the decision-based estimator is equal to GREG estimator for the union of two sample sets, that is, for $S = S_1 \cup S_2$. Otherwise, the decision-based estimator is the sum of two separate GREG estimators of stratum totals, that is,

$$\hat{t}_{y,dec} = \begin{cases} \hat{t}_{y,reg} & \text{if } \mathrm{H}_0 \text{ is accepted} \\ \sum_{h=1}^{2} \hat{t}_{y,reg}^{\,h} & \text{if } \mathrm{H}_0 \text{ is rejected} \end{cases} \qquad (6)$$

where $\hat{t}_{y,reg}$ denotes the GREG estimator from equation (2) for the combined stratum S, while $\hat{t}_{y,reg}^{\,h}$ denotes the GREG estimator from (2) for the total of substratum h units based on S_h sample data.

Table 1 shows real data examples of the hypothesis test statistics and decisions in the hypothesis test of $H_0 : b_1 = b_2$. The **Table** contains 18 tests for 3 variables (full-time payroll, full-time employment, and part-time payroll), 4 states (1=Alabama, 5=California, 39=Pennsylvania, and 50=Wisconsin), and 2 government types (30=sub-counties and 40=special districts).

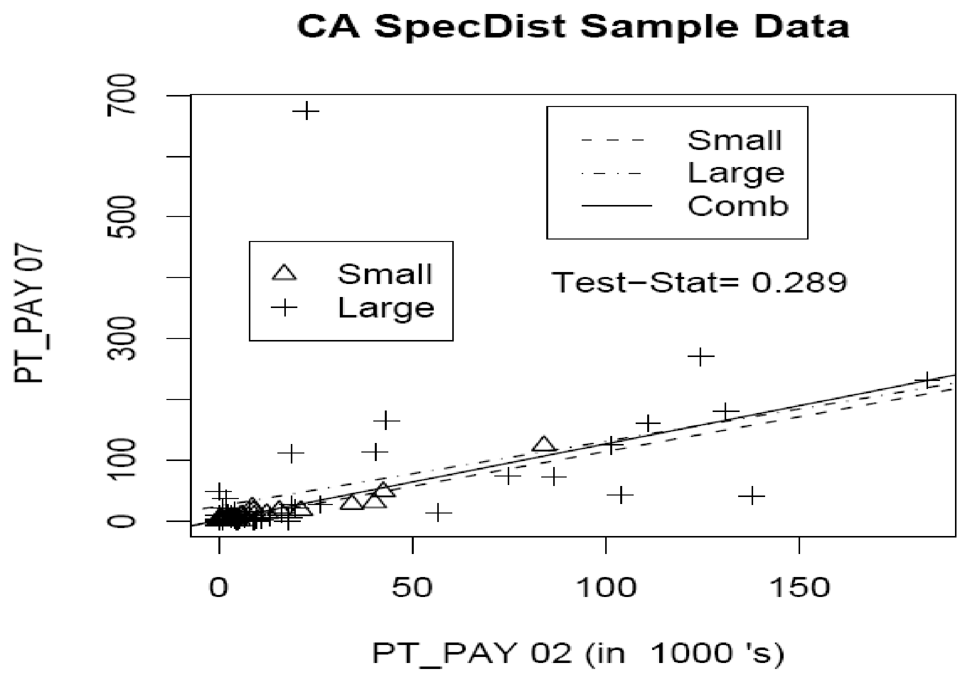
Table 1: Test results for decision-based using 2007 employment census data

	FT_Pay		FT_Emp		PT_Pay	
(State,Type)	Test-Stat	Decision	Test-Stat	Decision	Test-Stat	Decision
(1,30)	2.06	Reject	2.04	Reject	3.62	Reject
(5,40)	0.98	Accept	1.02	Accept	0.29	Accept
(39,30)	0.54	Accept	0.62	Accept	0.08	Accept
(39,40)	0.24	Accept	0.65	Accept	1.09	Accept
(50,30)	0.57	Accept	0.85	Accept	2.11	Reject
(50,40)	1.33	Accept	0.85	Accept	2.52	Reject

Source: U.S. Census Bureau. 2007 Census of Governments: Employment

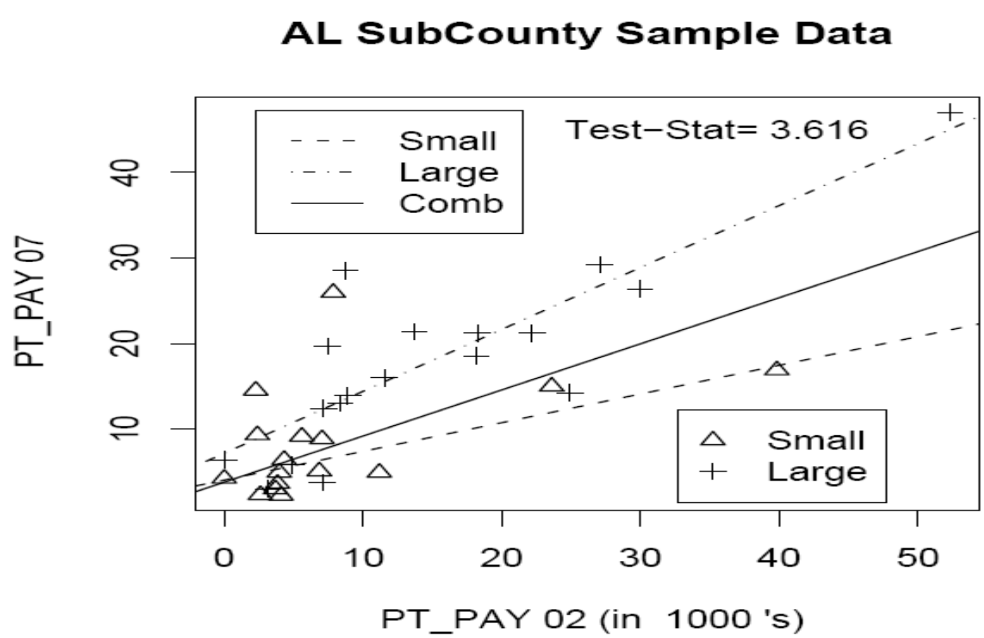
We display two real examples with 2007 Census of Governments: Employment data. **Figure 3** shows three sample-weighted linear regressions fitted to the small-unit and large-unit substrata and the combined stratum, for California Special Districts with part-time payroll as the response variable. The test statistic is 0.289, which is less than 1.96, so that the decision-based estimator is the GREG estimator (2) for the combined stratum.

Figure 3: Decision-based estimation for California special districts



Source: U.S. Census Bureau. 2002 and 2007 Census of Governments: Employment

Figure 4: Decision-based estimation for cities and towns in Alabama



Source: U.S. Census Bureau. 2002 and 2007 Census of Governments: Employment

Figure 4 similarly shows substratum and combined-stratum sample-weighted linear regressions fitted for Alabama sub-county government units, also with part-time payroll as y-variable. But in this second example, the test statistic is 3.616, much larger than 1.96. Thus, the decision-based estimator is the sum of the GREG estimators for the small- and large-unit substrata.

3. Variance Estimation for the Decision-Based Estimator

In order to compute the variance estimator for survey estimates based on unequal probability sampling, whenever possible we apply Horvitz-Thompson variance estimators such as the classical Sen-Yates-Grundy estimator. Such estimators rely on design joint inclusion probabilities, which can be laborious to specify for PPS selection methods without replacement. For at least one commonly used PPS without-replacement design, Vijayan's (1968) extension of a method of Hanurav (1967), this is now easy in SAS: the Vijayan-Hanurav design is the default PPS method in the new SAS sample selection procedure, PROC SURVEYSELECT, which is now used in selecting PPS samples within state-by-government type strata in the Annual Survey of Public Employment and Payroll. The formula-based joint inclusion probabilities can be computed directly and stored using options in PROC SURVEYSELECT.

However, a convenient and generally accurate approximate variance formula which avoids the need for joint inclusion probabilities is the PPS *with replacement* (PPSWR) variance estimator

$$\hat{V}(\hat{t}_y) = \frac{n}{n-1}\sum_{i\in S}(z_i - \bar{z})^2 \tag{7}$$

where $z_i = \dfrac{y_i}{\pi_i}$ and $\bar{z} = \dfrac{1}{n}\sum_{i\in S} z_i$. The PPSWR variance estimator usually overestimates the true variance slightly.

We adapt the PPSWR variance formula to cover the GREG estimator (2) based upon a PPS sample, as follows. The PPSWR approximation for the theoretical variance (within a single stratum) is

$$V(\hat{t}_{y,reg}) = \sum_{i\in U}\frac{e_i^2}{p_i} \tag{8}$$

where p_i is the single-draw probability of selecting a sample unit i and $\pi_i = np_i$ for a PPS sample of size n, and e_i is the population regression residual. The variance is estimated by the quantity (incorporating an estimated regression residual)

$$\hat{V}(\hat{t}_{y,reg}) = \frac{n}{n-1}\sum_{i\in S}\frac{\hat{e}_i^2}{\pi_i^2} \tag{9}$$

The variance of $\hat{b}$, the GREG slope estimator in equation (4), is needed in each substratum for the hypothesis test statistic of equal substratum regression slopes. To estimate this variance, we again adapt the PPSWR approximation to express the theoretical variance of $\hat{b}$ as

$$V(\widehat{b}) = \frac{1}{\sum_{k \in S} \frac{1}{\pi_k} (x_k - \widehat{t}_{x,\pi} / \widehat{N})^2} \frac{1}{n} \sum_{i \in U} p_i (\frac{x_i - \widehat{t}_{x,\pi} / \widehat{N}}{p_i})^2 \tag{10}$$

and to estimate it by

$$\widehat{V}(\widehat{b}) = \frac{n}{n-1} \sum_{i \in S} [\frac{\widehat{e}_i (x_i - \widehat{t}_{x,\pi} / \widehat{N})}{\pi_i}]^2 \Bigg/ [\sum_{i \in S} \frac{(x_i - \widehat{t}_{x,\pi} / \widehat{N})^2}{\pi_i}]^2 \tag{11}$$

Based upon the estimator (9), we produce a naive variance estimator $\widehat{V}_N$ for the decision-based stratified GREG estimator. Using sample data from two substrata as in (6), the decision-based estimator is either the GREG estimator on the stratum formed as the union S of substrata when $H_0 : b_1 = b_2$ is accepted, or is the sum of separate substratum GREG estimators when this hypothesis is rejected. In the case of accepting the null hypothesis, $\widehat{V}_N$ is the whole-stratum variance estimator (9); when the hypothesis is rejected, $\widehat{V}_N$ is defined as the sum of the separate substratum estimators (9). This is a naïve variance estimator, in the sense that it ignores the randomness inherent in the hypothesis-test-based decision, and treats that decision as though it were known in advance. Such a variance estimator might be expected to work well in settings where almost all of the P values for estimated differences of high- and low-z substratum slopes are very large or very small. (This requirement is only that each state-by-type hypothesis test is extremely decisive, not that all the decisions, to combine or not to combine high and low-z substrata, come out the same way.) However, if many of the pooling decisions are not extremely clear, then one might expect that the extra variability involved in the pooling decisions ought to inflate the variances of the estimated totals beyond what these naïve estimators say.

We now proceed to develop and validate a more sophisticated estimator of the variance of the stratum-wise regression estimator following decision-based pooling, using bootstrap and Monte Carlo methods, and to examine the extent to which it exceeds the naïve estimators of variance.

4. Bootstrap and Monte Carlo Variance Estimation

In statistical estimation problems based on complicated or multistage decisions, bootstrap methods have become an essential and theoretically supported tool for variance estimation (Shao and Tu, 1995). To some extent, the bootstrap methods have also been shown to generate valid variances and sampling distributions for estimators based on surveys, at least under stratified SRS survey plans. In the present context, where PPS sampling is used within (state by government-type) strata, the bootstrap theory would apply in large-sample settings in which almost all or almost none of the bootstrap-resampled hypothesis tests for equality of substratum slopes would reject. However, in such settings the Naive variance estimator described above also has large-sample theoretical justification. Moreover, the attractive feature of the decision-based estimation strategy is to allow pooling in those strata where the test does not reject, and the proportion of these is in practice clearly different from 0 and 1. Data analyses in a selection of six state-by-type strata are summarized in **Table 1**, showing that the hypothesis testing decisions will in fact not always be extremely clear, and will be sometimes to pool the small- and large-unit substrata and sometimes not.

Bootstrap methods generally involve drawing many replicate with-replacement samples with equal probability and fixed size from each stratum of a survey dataset and re-analyzing each of these bootstrap survey-samples using exactly the same steps which were used in the estimation method under study. In our setting, consider the case of a single state-by-type stratum in which a PPS sample of size n is drawn, and in which the stratum is initially decomposed further into sub-strata of small and large units, with respective sample sizes n_1 and n_2. Bootstrap samples of respective sizes n_1 and n_2 are drawn with equal probability from n_1 and n_2 sampled governmental units in the original substrata. The bootstrapped attribute samples x_i^*, y_i^* in the two substrata are used to estimate simple weighted linear regression slopes $\widehat{b}_j^*$ for j = 1 and 2, and the hypothesis test of equality of slopes is conducted by respectively rejecting or accepting according as

$$\frac{|\,\widehat{b}_1^* - \widehat{b}_2^*\,|}{[SE(\widehat{b}_1^*)^2 + SE(\widehat{b}_2^*)^2]^{1/2}} \geq 1.96 \qquad (12)$$

where the standard errors in the denominator are estimated as in (11). In case of acceptance, denoted $T^* = 1$, the bootstrap estimator $\widehat{t}_{y,dec}^*$ is the whole-stratum y-total regression estimator, and otherwise when $T^* = 2$, the decision-based bootstrap estimator is the sum of the two substratum estimators. The bootstrap naïve variance estimator $\widehat{V}_N^*$ is calculated from the bootstrap sample, using the PPS (with-replacement) variance formula either based on 1 or 2 substrata (according to whether $T^* = 1$ or 2) as though the hypothesis test outcome were known a priori. This bootstrap replication is repeated B times, producing triples $\widehat{t}_{y,dec}^{*b}$, $\widehat{V}_N^{*b}$, and T^{*b} for b = 1, 2, …., B.

The bootstrap variance estimator for the decision-based y-total estimator $\widehat{t}_{y,dec}$ is then the sample variance of the B bootstrap replicates $\widehat{t}_{y,dec}^{*b}$ for b = 1, 2, …, B , and our objective is to compare it with the naïve estimators $\widehat{V}_N(\widehat{t}_{y,dec})$ and the sample mean of $\widehat{V}_N^{*b}$ for b = 1, 2, …, B.

As mentioned above, we expect the bootstrap variance estimators to be roughly valid and roughly equal to the naïve estimators in situations where the bootstrapped rejection indicators are nearly all 1 or nearly all 0. However, in settings where the bootstrap proportion of rejections is well away from 0 or 1, we should expect differences. Moreover, such cases are similar to bootstrap examples where Shao (1994) has shown (in a similar but simpler setting of estimation of a mean following a hypothesis test) that bootstrap sample sizes should be much smaller than their respective substratum sizes n_k in order to provide consistent distribution and variance estimates for bootstrapped statistics dependent on the outcome of a hypothesis test. However, since the reason for the inconsistency was the mixture distribution for the final estimator resulting from the proportion 0.05 of cases where rejection occurs, we viewed the inconsistency as possibly a subtle or second-order effect and experimented only with bootstrap substratum samples of the same sizes as the original samples. We describe next a simulation study performed to explore the performance of bootstrap variance estimation in this setting.

We use Monte Carlo simulated artificial superpopulations to check the validity of bootstrap and naïve variance estimates. In this validation experiment, we begin by simulating an artificial superpopulation of two substrata of respective sizes N_1, N_2 and hold that superpopulation fixed. We then draw independently, successively over a large number R of Monte Carlo replications, PPSWR subsamples of n_1 and n_2 elements from the two substrata. Within each such sample indexed by r=1,…,R, we compute the decision-based estimator and its naïve variance estimator, and perform B bootstrap replications.

5. Data Simulation and Numerical Results

We describe in this Section the results of a simulation study to compare bootstrap and naïve variance estimators for the decision-based y-total estimators, and to assess the quality of the bootstrap variance estimator (always based on bootstrap samples of the same sizes n_1 and n_2 as the original samples drawn.

In this simulation, the attributes (x_i, y_i) are jointly generated from fixed and known probability distributions, according to the following steps. For fixed N_1, N_2, we generate $N = N_1 + N_2$ independent identically distributed variates x_i fro m a Gamma (α, β) distribution with specified mean and variance. Then the indices with the N_1 lowest x_i values, say all i for which x_i is less than c, define substratum 1, with the remaining indices in substratum 2. Let U_1 and U_2, respectively denote the substratum index sets. The frame-substratum survey totals $t_{x,h}$ for h=1,2, are fixed in each superpopulation to be the respective x_i totals over all indices $i \in U_h$.

We generate $e_{hi} \sim N(0, \sigma_h^2)$, independently of each other and the x_i values, for the N_h indices $i \in U_h$, for h = 1 and 2. Next, we generate y_i according to the rule

$$y_i = \begin{cases} a_3 x_i^2 + a_2 x_i + a_1 + e_{1i} & i \in U_1 \\ a_3 x_i^2 + a_2 x_i + a_1 + d(x_i - c) + e_{2i} & i \in U_2 \end{cases} \tag{13}$$

The purpose of this model definition is to allow patterns of expected (x, y) dependence encompassing equal lines, lines with differing slopes (but same intercept), and quadratic curves.

Finally, the substratum samples of size n_1, n_2 are drawn PPSWR within each substratum with size-variable equal to the same x_i values, i.e., the PPS inclusion probabilities within each substratum are taken proportional to the x_i values. The PPSWR sample weights within substratum h are $1/\pi_{hi} = w_{hi} = \sum_{j \in U_h} x_j \Big/ n_h x_i, i \in U_h$.

Table 2 lists the parameters $(a_1, a_2, a_3, d, \sigma_1^2, \sigma_2^2)$ used in our data simulations, along with the frame and sample sizes. All of the x_i values in these experiments were simulated as Gamma (9, 3/2), with these parameters chosen so that the x_i have mean 6 and standard deviation 2. Note that although the N_1, N_2 values chosen here result in two substrata of relatively balanced size, the low-size substratum is generally much larger in the Governments survey within each state-by-type stratum, so that further simulations will be needed in the future to understand the impact of these results.

Table 2: Data Simulation Parameter Table

Examples	a_1	a_2	a_3	D	σ1	σ2	n1	n2	N1	N2
1	0	2.0	0.2	0.0	3	3	40	60	1500	1200
2	0	2.0	0.0	0.2	3	3	40	60	1500	1200
3	0	2.0	0.0	0.4	3	3	40	60	1500	1200
4	0	2.0	0.0	0.6	3	3	40	60	1500	1200
5	0	2.0	0.0	0.6	4	4	40	60	1500	1200
6	0	2.0	0.0	0.8	4	4	40	60	1500	1200
7	0	2.0	-0.1	0.8	4	4	40	60	1500	1200
8	0	2.0	0.2	0.0	3	3	20	30	1500	1200
9	0	2.0	0.0	0.2	3	3	20	30	1500	1200
10	0	2.0	0.0	0.4	3	3	20	30	1500	1200
11	0	2.0	0.0	0.6	3	3	20	30	1500	1200
12	0	2.0	0.0	0.6	4	4	20	30	1500	1200
13	0	2.0	0.0	0.8	4	4	20	30	1500	1200
14	0	2.0	-0.1	0.8	4	4	20	30	1500	1200
15	0	2.0	0.0	0.0	3	3	30	45	1500	1200
16	0	2.0	0.0	0.0	4	4	30	45	1500	1200
17	0	1.6	0.0	0.0	3	3	30	45	1500	1200
18	0	1.6	0.0	0.0	3	4	30	45	1500	1200
19	0	1.6	0.0	0.0	3	5	30	45	1500	1200
20	0	1.3	0.0	0.0	3	5	30	45	1500	1200

Source: Parameters used in simulating data, for illustrative purposes only

Now, we define null hypothesis reject rates for decision-based methods. Prej.MC is the proportion of rejections in the hypothesis test for equality of slopes in the Monte Carlo method, and Prej.Boot is the proportion of rejections in the hypothesis test for equality of slopes over all R*B bootstrap-by-superpopulation replications.

In the tabulated simulation results, we exhibit naïve and empirical Monte Carlo and Bootstrap variance estimators, as well as Mean Square Errors for both the decision-based survey estimates (DEC.MSE) and the estimates which always use 2 substrata (2str.MSE). The naïve variance estimators for Monte Carlo and Bootstrap are averages over replicated naïve variance estimators, respectively, that is, $V_{MC,naiv} = \frac{1}{R}\sum_{r=1}^{R} V_{naiv}^{r}$ where V_{naiv}^{r} is the naïve variance estimator from (9) for the r'th Monte Carlo replication, and $V_{Boot,naiv} = \frac{1}{RB}\sum_{r=1}^{R}\sum_{b=1}^{B} V_{naiv}^{rb}$ where V_{naiv}^{rb} is the naïve variance estimator for the b'th bootstrap replication within the r'th Monte Carlo replication. The empirical variance estimators for Monte Carlo and Bootstrap are respectively the sample variances of the decision-based estimates and the averages over the Monte Carlo replicated Bootstrap sample variances.

Table 3: **Simulation results with R=500 and B=60. Columns 2-3 contain rejection probabilities, columns 4-7 square roots of estimated variances, and 8-9 empirical MSE's.**

Ex.	Prej. MC	Prej. Boot	MC. Emp	MC. Naiv	Boot. Emp	Boot. Naiv	DEC. MSE	2str. MSE
1	.796	.719	991.8	867.9	863.6	846.9	832904	819736
2	.098	.231	920.6	873.2	871.4	856.4	846843	857654
3	.126	.277	908.3	868.6	903.2	847.0	826142	845332
4	.258	.333	880.9	874.7	862.8	850.6	777871	779790
5	.144	.249	1159.5	1139.0	1192.1	1111.4	1346545	1351290
6	.258	.339	1173.5	1144.1	1179.1	1113.7	1374466	1401604
7	.088	.217	1167.7	1148.4	1165.3	1126.7	1361384	1397779
8	.582	.601	1288.2	1209.1	1229.4	1149.8	1656195	1656324
9	.108	.236	1174.0	1169.4	1291.3	1118.7	1376302	1432493
10	.164	.283	1377.3	1186.0	1211.2	1129.8	1907303	1992299
11	.188	.301	1339.1	1179.6	1236.7	1128.6	1791618	1890654
12	.108	.261	1612.1	1609.8	1676.6	1529.5	2594122	2648593
13	.212	.322	1654.3	1566.3	1668.9	1497.1	2736425	2762037
14	.116	.254	1559.8	1564.5	1642.6	1490.9	2456142	2563424
15	.100	.230	982.3	961.4	973.4	924.4	963009	985644
16	.120	.264	1319.8	1308.3	1307.6	1264.3	1738460	1799603
17	.102	.243	984.0	988.1	995.2	959.9	966444	999423
18	.074	.220	1105.0	1106.0	1134.6	1076.1	1219898	1252016
19	.078	.217	1301.8	1236.2	1211.6	1203.4	1696521	1712228
20	.102	.235	1305.1	1256.5	1275.1	1223.4	1700264	1739345

Source: Data are simulated for illustrative purposes only

The formulas are $V_{MC,Emp} = \frac{1}{R-1}\sum_{r=1}^{R}(\hat{t}_{y,dec}^{r} - \bar{t}_{y,dec})^2$ and $V_{Boot,Emp} = \frac{1}{R}\sum_{r=1}^{R} S_r^2$, where $\bar{t}_{y,dec} = \frac{1}{R}\sum_{r=1}^{R}\hat{t}_{y,dec}^{r}$ and S_r^2 is the sample variance of $\{\hat{t}_{y,dec}^{rb}, b = 1,\cdots B\}$. Of these variances, MC.Emp best estimates the true variance of $\hat{t}_{y,dec}$ for the simulated superpopulation and sample design, and Boot.Emp estimates the average of variance estimated by the Bootstrap method .

The tentative conclusions from this simulation study are:

(i) that the bootstrap estimate of the probability of rejecting the null hypothesis of equal substratum slopes can be quite different from the true probability;

(ii) that the Naïve estimator of standard error of the decision-based estimator is generally slightly less than the actual standard error;

(iii) that the Bootstrap estimator of standard error is not reliably close to the true standard error (the MC.Emp column); and

(iv) that the mean-squared error for the decision-based estimator is generally only slightly less than that for the two-substratum estimator, but does seem to be a few percent better for a broad range of parameter combinations.

In (iii), although we see discrepancies of no more than about 10 percent in these reported simulations, other preliminary simulations with small- and large-unit substrata or more unequal size and with longer-tailed distributions for y can make the discrepancies much greater.

6. Future Research

We plan to investigate the large-sample asymptotic properties of the bootstrap in this setting. With reference to Shao and Tu (1995), and to simulations not reported here using sample sizes n_1, n_2 too large to be practical in the context of Governments Division surveys, the existing bootstrap theory works well in a setting with probability proportional-to-size with-replacement sampling except under the null hypothesis of equal regression slopes.

To supplement the Bootstrap variance estimation techniques, random-groups balanced repeated replication (BRR) is worth exploring as an alternative way of estimating variances for the decision-based survey estimators described in this paper.

When we applied the decision-based estimators to some real Governments Division survey state-by-type strata, and to simulated data with large N_1 and small N_2 , fairly small n_1, n_2, and heavy-tailed distribution for y (resulting frequently in some high-leverage points y_i), we found that the bootstrap variance estimator for the decision-based estimator can be several times larger than the Naïve and actual variances. Thus, we believe that an important issue for further research is the development of robust or outlier-resistant regression techniques in defining decision-based estimators and their variance estimators.

Acknowledgements

We thank Dr. Jun Shao from the University of Wisconsin at Madison and the Statistical Research Division of the Census Bureau for contributing ideas about bootstrap variance estimation and Casey Corcoran from the Governments Division of the Census Bureau for help with data analysis and SAS programming. Also, we are indebted to our reviewers, Lisa Blumerman from the Governments Division of the Census Bureau and Dr. Patrick Cantwell from the Decennial Statistical Studies Division of the Census Bureau for their helpful suggestions, which improved the original paper.

References

Barth, J., Cheng, Y., and Hogue, C. (2009), Reducing the Public Employment Survey Sample Size, JSM 2009 Proceeding.

Cheng, Y., Corcoran, C., Barth, J. and Hogue, C. (2009), An Estimation Procedure for the New Public Employment Survey Design**,** JSM 2009 Proceeding.

Cochran, W.G. (1977), Sampling Techniques. Third Edition. New York: John Wiley & Sons, Inc.

Deville, J-C. and Sarndal, C-E. (1992), Calibration Estimators in Survey Sampling, Journal of the American Statistical Association, Volume 87, Number 418, 376-382

Hanurav, T.V. (1967), Optimum utilization of auxiliary information: πps sampling of two units from a stratum, Journal of the Royal Statistical Society, Series B, Volume 29, 374-391

Sarndal, C.-E., Swensson, B., and Wretman, J. (1992), Model Assisted Survey Sampling, Springer-Verlag.

Shao, J. (1994), Bootstrap Sample Size in Nonregular Cases, Proceedings of the American Mathematical Society, Volume 122, Number 4, 1251-1262

Shao, J. and Tu, D. (1995), The Jackknife and Bootstrap, Springer-Verlag.

Vijayan, K. (1968), An Exact πps Sampling Scheme-Generalization of a Method of Hanurav, Journal of the Royal Statistical Society, Series B, Volume 30, Number 3, 556-566

Wolter, K. (2007), Introduction to Variance Estimation, Second Edition, Springer-Verlag.